1.0 Successful Information Management for EMPACT Projects

1.1 Overview

Rapid advances in technology have made it possible for millions of people to obtain access to near real-time information and data via the World Wide Web. In 1996, President Clinton directed EPA to create a new program to take advantage of these and other capabilities to bring useful environmental information to the American public in near real time. In response to the President's directive and the opportunities created by the Web, EPA launched the Environmental Monitoring for Public Access and Community Tracking (EMPACT) program.

EMPACT is working to achieve an aggressive, forward-thinking vision:

> To provide time-relevant, environmental information to a broad spectrum of the American public . . . in a framework that supports national consistency, but also encourages leveraging new and innovative measurement, monitoring and modeling technologies and resources through partnerships to meet the decision-making needs of the public.

The goal of EMPACT to provide members of the public access to clearly communicated, timely, and accurate environmental and public health information, so that they can make informed decisions about their day-to-day lives.

1.2 Purpose of the Handbook

The purpose of this Handbook is to provide guidance on managing EMPACT projects, particularly with respect to data management and documentation, and on disseminating environmental information to the public. In particular, the Handbook does the following:

- summarizes best management practices for EMPACT projects;
- describes how EMPACT project data should be documented;
- provides guidance on format and content of EMPACT Web sites;

- highlights how other communications tools besides the World Wide Web can be used to

disseminate EMPACT information;
- provides requirements for using metadata; and
- provides examples of EMPACT projects that are already underway.

The intended audience of this document includes project managers, data managers, quality assurance/quality control managers, information specialists, Webmasters, and EMPACT regional coordinators.

1.3 Key EMPACT Information Management Activities

There are five major kinds of EMPACT information management activities:

1. Planning the project
2. Gathering the data
3. Managing the data
4. Disseminating the data
5. Evaluating results.

These activities occur in sequence, with the understanding that each activity can be adjusted to reflect changing conditions and project results. Other crosscutting activities should occur throughout the lifetime of a project, as follows:

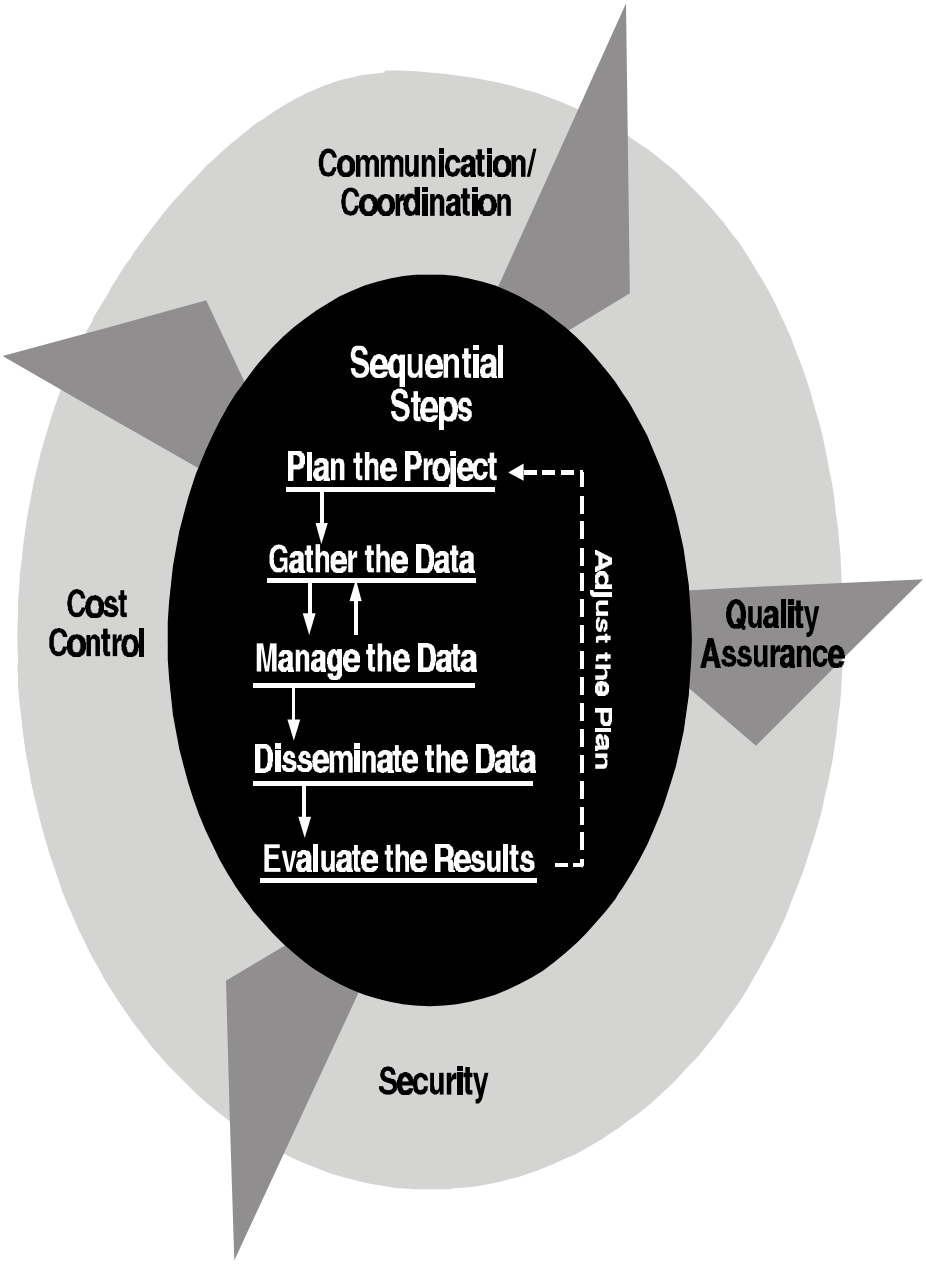
- communication and coordination,
- quality assurance,
- cost control, and
- security.

Figure 1-1 illustrates the sequence of project management activities and cross cutting activities. This general model should be applied to all EMPACT projects.

This Handbook focuses on the information management aspects of EMPACT projects (managing and disseminating data). It does not address the other aspects of project management (planning, data gathering, and evaluation).

The EMPACT Information Management Checklist (following the Foreword) summarizes the key tasks that must be carried out in managing data and disseminating data.

Figure 1-1: Key Project Management Activities



1.4 Best Management Practices for EMPACT Data and Information

In creating this Handbook, EPA identified best management practices among a broad range of projects. As described below, those practices are common to all successful time-relevant environmental monitoring projects. They should be tailored to address individual project circumstances. The best management practices described in this section are based on current technology. Technology changes rapidly but these best management practices can be expected to remain valid.

1.4.1 Use a Common Data Format

Many EMPACT projects have faced a major challenge in converting data collected from monitoring devices to a format that can be incorporated into a database. The ASCII comma delimited format is an effective tool for transferring data and many EMPACT projects have used to create computer readable data files. It can be read by almost any machine or software and is easily transferrable. The risk of this format is that missing fields can destroy the value of the entire data set. As long as there are no missing fields in the data, the ASCII comma delimited format is effective for building data sets from data supplied by monitoring devices.

1.4.2 Provide Data by an Open Access System

Information management systems developed for EMPACT projects must provide open access formats that allow the public to make secondary uses of project data. An open system is one that adheres to a publicly known and sometimes standard set of interfaces so that anyone using them can also use any other system that adheres to that standard. To achieve this "interoperability," open systems should be built so that neither the design nor the implementation of the system would lock the application into a particular vendor hardware or software platform. Open systems should have the following features:

- compliance with industry standards for programming, communications, networking, system management, presentation, and interfaces;
- portability of applications across systems;
- scaleability of applications performance and throughput; and
- interoperability across systems.

Open databases are an important part of open systems. Open databases follow two types of standards, one that specifies a query language and one that specifies "interoperability" of tools.

SQL (standard query language) is a standard interactive and programming language for requesting information from and updating a database and is the most common means of addressing the first

standard relating to query language. SQL statements are used both for interactive queries for information from a relational database and for gathering data for reports.

Open Database Connectivity (ODBC) is a standard or open application programming interface (API) for accessing a database. The goal of ODBC is to make it possible to access any data from any application, regardless of which database management system is handling the data. ODBC accomplishes this by inserting a middle layer, called a database driver, between an application and the database management system. The purpose of this layer is to translate the application's data queries into commands that the system understands. For this to work, both the application and the system must be ODBC-compliant; that is, the application must be able to issue ODBC commands and the system must be capable of responding to them. ODBC is based on and closely aligned with SQL. ODBC handles the SQL request and converts it into a request the individual database system understands.

Open access databases make it easier for secondary facilitate to obtain and process data for their own purposes. The secondary user does not have to actually own the data, since it is already being supplied by another source. That data can then be interpreted on its own or integrated with data sets from other sources. Examples of open access databases include Oracle, SQL Server, DB2, Informix, Microsoft Access, and Sybase.

1.4.3 Document the Project

Each EMPACT project must be thoroughly documented, so that all parties involved in the project know what project information management systems are designed to do and how they are built and operated. Careful documentation also allows projects to be compared to one another and new projects to learn from more established ones. EMPACT projects should be documents in four ways:

- overall project documentation;
- data set documentation;
- data element documentation; and
- database documentation.

Project documentation is discussed in more detail in section 2.1.4 and on the EMPACT Web site (*http://www.epa.gov/empact)*.

1.4.4 Establish Back Up and Archiving Procedures

Each EMPACT project should establish standard procedures for data backup and archiving, system failure, and recovery.

Backup procedures describe a method involving both software and hardware to perform regular backups of project data. Their purpose is to ensure speedy recovery from data-related disasters.

Losses could result from such ordinary phenomena as an electrical storm or accidental overwrite to malicious destruction by hackers. The methods used will vary with the scale of the project and the criticality of the data loss**.**

Data sets should be archived to provide a historical record of data over the lifetime of the project. Procedures should be established for archiving methods, frequency, and media.

Failure procedures are performed after system failures. They would apply to projects that maintain their own computer servers. Procedures might include using a manual system or an alternative computer-based backup system.

Recovery procedures restore the system to normal operation following a system failure. They range from rebooting a computer to re-keying data from the period of system failure. Procedures should clearly document how a database is to be restored to a consistent state as well as how defective hardware and software are to be replaced.

1.4.5 Ensure Data Integrity

Ensuring data integrity involves addressing the vulnerability of the system to unauthorized access, data manipulation, theft, and environmental damage. Distributed systems, such as Web applications, are particularly vulnerable to security violations. Most threats are internal and are caused by human error and omissions. Some key potential threats include the following.

Inappropriate/inaccurate information - Project workers can supply inappropriate or inaccurate information accidentally or on purpose. This threat is usually created by human error in combination with undefined or weak internal quality assurance/quality control processes.

Compromised information integrity - Documents or data can be accidentally or intentionally modified.

Compromised server availability - Human or environmental factors may result in temporary slowdown or complete loss of public access to information provided by the servers. Programmers may disrupt server function by poor crafting of applications. Servers may also be sabotaged through intentional or unintentional introduction of computer viruses. Environmental hazards include electrical power failures/disturbances, hardware failures, fire/water damage, and air-conditioning malfunctions.

Security must be commensurate with the risk and the magnitude of harm resulting from the loss or damage to information or equipment. Protective measures include database replications, server backups, and password protection. Securing the site where servers are located is also important**.**

1.4.6 Present Data in an Understandable Format

EMPACT projects should present information in a way that is suitable for the intended audience. The following guidelines have been helpful in designing effective approaches to communication and should be considered in designing EMPACT projects:

Establish a Context for Presenting the Data to the Public - In addition to providing interpretation for time-relevant data, it is often important to place the information in context for a community or audience. Methods for providing context include displaying data in a geographic context, combining the new data source with existing collections of monitored data, or aggregating the data collected to demonstrate a trend or ongoing view of the environmental conditions. Whatever method is selected, projects must try and respond to the needs of their users and present the data in ways that make the data relevant and useful for them.

Format Data For Easy Interpretation - Environmental data can be interpreted in many different ways. The conclusion a hydrologist draws from a set of water quality data may differ greatly from that of a layperson, yet both may be valid. Data should always be presented in a format that is easy to understand and not subject to misinterpretation. Any interpretation of data provided by the project should be clear to the intended.

Be Responsive to the Users of Your Data - Although, each EMPACT project is directed at a specific audience, it is not always possible to know in advance who will find project information useful. If the audience has changed, or unexpected audiences are using project data, projects should change as needed to accommodate new audiences. To keep abreast of changing user needs, EMPACT projects should build in ways for users to provide feedback on the projects. For example, the USGS Stream-Gaging Program was developed to predict floods and to be used by forecasters, emergency response teams, and the like. The program and its Web site, however, have become a popular destination for canoeists, white-water rafters, and other water adventure seekers. The USGS never imagined these groups of people would be using their data and have since accommodated their needs.

By tailoring a project to a certain audience, the project risks excluding potential users. Indeed, a certain amount of tailoring of information is always necessary to give the project definition and scope. Nonetheless, known secondary users should not be excluded.

Section 3 provides additional examples and guidance for presenting data and information effectively.

1.4.7 Ensure Data Quality

All EMPACT projects are required to develop a quality assurance/quality control (QA/QC) plan. In developing these plans, projects should consider three key items: quality assurance needs analysis, systems development planning, and audits and testing. EPA provides extensive guidance on quality assurance planning for information management projects at *http://www.epa.gov/r10earth/offices/oea/epaqag5.pdf*

Quality Assurance Needs Analysis The type of QA effort needed depends on the qualitative and quantitative criteria that the data must meet and on the complexity and magnitude of the project. Other specific concerns such as security and system performance also help define the QA program requirements. Only by establishing the ultimate needs and objectives for the data quality in the early planning stages can appropriate decisions be made to guide the system development process to a successful conclusion.

Systems Development Planning Proper planning, execution, and QA protocols are vital to the success of projects involving information systems development, software development or computer data processing. The project management team should work closely with the responsible QA staff to implement a program that best suit the needs of the individual project. A few of the issues addressed include the level of documentation required, schedule, personnel assignments, and change control. The following section describes a commonly used planning framework and associated documentation that is based on the widely recognized software or system -development life cycle.

Audits and Testing As with any project involving generation or handling of environmental data, audits can be used to verify that goals and objectives are being met. Audits of the Information System Development Process, audits of security and data verification audits may be particularly helpful when conducted by personnel outside the immediate project team. Security audits by knowledgeable experts be valuable when data confidentiality and prevention of tampering are important issues. Data verification audits can be conducted using a known data set. Such a data set might be developed by an end-user or an outside expert to verify that the information system produces the expected results.

The purpose of testing is not to simply detect errors but also to verify that the completed software meets user requirements. In designing any test, the “correct” or “acceptable” outputs should be known in advance, if possible. Testing should be planned in an orderly way and documented.

Even when QA procedures are performed on time-relevant data, the fact remains that the data usually has not been completely scrutinized. Although the QA plan should be thorough enough to detect major quality issues, the nature of time-relevant data and the relatively short turn around time available for quality assurance frequently does not allow for exhaustive QA measures. Thus,

whenever time relevant-data is presented, users should be cautioned that the data is provisional.

1.5 EMPACT Information Management Plan Requirements

Each project is required to develop an Information Management Plan that documents its approach to data collection, storage, retrieval, delivery, and communication and procedures for data quality control and security. The need for a quality preliminary Information Management Plan is critical to ensure the effective allocation of resources and delivery of EMPACT data in a timely and useful manner. All data collected must be made available in electronic format and stored in a computer.

The Information Management Plan must include the following elements:

✔ **Name(s) of the Data Owners**

Provide the primary contact information of the individual and organization that has overall information management responsibility and authority for your project. Please provide their name, address, telephone and email address if possible. Also, please provide contacts for organization(s) of key partners collecting and/or managing the data.

✔ **Description of the Data Flow Process**

Explain the key points in the information flow pipeline data collection through public access. This can be achieved by use of a flow chart or diagram depicting key information components. Where applicable, be sure to illustrate components of the process that involve relationships with multiple partners. (e.g. academic institutions, vendors, Internet service providers).

✔ **Description of the Data Collection Methods**

This section should summarize data measurement and collection methods. Define the sampling frequency of the environmental parameter(s) to be measured and monitored. Describe any emerging measurement technologies used and/or how existing systems of environmental monitoring will be augmented / upgraded to provide the proposed data.

✔ **Description of the Data Storage and Retrieval System (Hardware and Software systems)**

Describe the hardware and software technologies (e.g. Unix, Win, Mac/database system) that will be used for data management and processing systems. Describe the components

of project architecture (e.g. client-server or mainframe, single or multi-sever). How will the project provide metadata search capabilities, if any? Describe the archival components or storage capabilities of your plan. Address data management responsibilities among the project's partners to ensure data documentation and data standardization.

✔ **Description of the Data Delivery System**

The key goal of EMPACT is to provide the public time-relevant information in a format they can understand. Describe specific technologies to deliver data to the user interface. At the very minimum, an World Wide Web home page for describing the program and posting project data should be established. Any other technologies that the project also intends to use (e.g. kiosk, telephone recording, television, radio, newspapers) should also be described.

List how products will be displayed for your Web site from the data collected. For example, will the information be in the form of:

- graphical charts;
- interactive or dynamically generated graphics;
- textual data;
- digital video or sound; or
- geospatial mapping.

What tools will the project use to provide the final product?

- particular server software?
- applications (non-database) used standalone or with a common gateway interface (CGI) (e.g. ArcView, Perl, Java)?
- any mechanisms used with the Web like an email alert that is also used for data delivery?

✔ **Information Management Plans Should be Limited to 6 pages**

Please keep this plan concise. It is not be necessary to go into complex technical details on how the project will use a particular piece of software or how hardware will be configured. The plan should provide the basic foundation for your project. One of the purposes of the plan is to provide information that can be shared with future EMPACT projects and to provide an information base for promoting sustainability of the projects after the EMPACT funding ends.